**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 1, 2022, p. 483-489
https://publishoa.com
ISSN: 1309-3452

# Early risk detection of depression from social media posts using Hierarchical Attention Networks

**Tamilarasan Ramasamy[1] and Dr. J. Jayanthi[2]**

Authors Affiliation
[1]Claritirics India Pvt Ltd, Chennai, Tamilnadu, India.
[1]Sona College of Technology, Salem 636005, Tamilnadu, India.
Authors
Email
[1]tamilarasan.r@gmail.com
[2] jayanthij@sonatech.ac.in

**Abstract.** Efficient mental health diagnosis is improving continuously, yet many cases go undetected. Early detection of depression can potentially prevent people from mental illness and live a better life. There are many ways to monitor depression in people; the most obvious one is to monitor the messages posted by people on social media platforms. In recent years, detecting early depression from social media posts has been a focused research area. In this paper, we use Hierarchical Attention Networks, a Deep Learning-based method to classify whether the users are depressed or not using their historical, social media posts.

## 1. INTRODUCTION

According to WHO, Mental health is a state of well-being in which an individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively, and is able to contribute to his or her community. Depression is one of the leading causes of disability. Suicide is the second leading cause of death among 15-29-year-olds. It is very common among people affected by communicable (e.g., HIV and TB) and non-communicable diseases (e.g., cancer and cardiovascular disease). Depression and anxiety disorders cost the global economy 1 trillion USD per year. It is estimated that there are 700,000 deaths every year from suicide globally, which is a leading cause of death in young people. Globally depression is the major reason for self-harm. Specifically, it leads to suicides in young adults across the globe [11].

Early detection of depression has several advantages, including safety and better health. Early warnings can help prevent people from getting into depression, self-harm, or eating disorders. Early detection technologies also help identify criminals when they start posting anti-social messages in internet forums. Many Machine Learning and Deep Learning based methods proposed in the recent times to analyze social media posts to flag whether the user is depressed. Machine learning based depression detection can bring in a different perspective in the field of Mental Health [1]. While depression and other mental illnesses may lead to social withdrawal and isolation, it was found that social media platforms are indeed increasingly used by affected individuals to connect with others, share experiences, and support each other [12][13].

## 2. RELATED WORK

Several machine learning and deep learning algorithms detected depression in social media posts. The following section provides some details on the ML and DL algorithms used in the early detection of depression.

### 2.1. Methods using Machine learning algorithms

For instance, Schwartz et al. proposed an approach to use a regression model based on Facebook data to predict multiple-granularity depression in individuals [16]. The authors proposed methods to find the degree of depression DDep. Instead of building a classifier system, the method continuously tracks the changes in DDep over time.

The authors estimated the degree of depression (DDep) as the average response to seven depression facet items nested within the larger Neuroticism item pool. For each item, users indicated how accurately short phrases described themselves (e.g., "often feel blue," "dislike me"; responses ranged from 1 = wildly inaccurate to 5 = very accurate).

Furthermore, many traditional machine learning algorithms were used in the shared task, automatic identification of content in mental health forums by the 2016 Computational Linguistics and Clinical Psychology Workshop.

For instance, Malmasi et al. used a Random Forest meta-classification approach on top of base classifiers [30]. The data for this shared task was taken from Reachout.com. Reachout.com is a support forum for online youth mental health. The authors used Lexical features such as Character n-grams, Word n-grams, Word skip-gram, Lemma n-gram, and Word representations. In addition to the lexical features, the syntactic features like POS tags, Dependencies, and production rules were used to build Random Forest-based classifiers.

In another paper, Brew used SVM with Radial Basis Function (RBF) kernel [31]. Brew also used the data from Reachout.com as part of the shared task. In this approach, features like unigrams and bigrams weighted with TFIDF, a feature representing author type, and the kudos that the users assigned to post were used to build SVM models.

## 2.2. Methods using Deep learning algorithms

In their paper, Zhang, Yipeng, et al. [6] proposed using Deep Learning models for chunk level classification for monitoring depression trends on Twitter data during Covid19. Their method focused on multi-channel CNN and bidirectional LSTM with context-aware attention, as described in Orabi, Ahmed Husseini, et al. [7]. They also used Glove vectors of dimension 300, Jeffrey Pennington et al. [8] in their approach.

In another study, Guntuku, Sharath Chandra, et al. [9] focused on linguistic representations of users' posts on Facebook and Twitter.

Instead of identifying the depression based on a specific post by the user, analyzing the messages posted over a period give more information about the user's behavior. Tuke, Curtis Murray, et al. [10] propose methods for extracting symptoms from social media posts. While the objective of this paper is not to focus on depression specifically, the method of using temporal information motivates the study of depression among the users over a period.

There were many shared tasks on depression detection in social media posts as part of the CLEF 2017 conference eRisk pilot task on early detection of depression; authors Trotzek, Marcel, et al. [11] collected data from reddit.com. The data contained chronological messages posted by the users and manually classified them under depressed or non-depressed classes. The methods used by Trotzek, Marcel et al.[11] uses features such as linguistics metadata, readability, emotions, and sentiments to train the neural network models. In their conclusion, the authors argue that the ERDEo metric for early detection tasks in detail is not a meaningful measure, specifically ERDE5, and proposed ERDE%0 as a new approach for better interpretation.

In their paper, Rao, Guozheng, et al. [14] proposed a hierarchical posts representations model named Multi-Gated LeakyReLU CNN (MGL-CNN) for identifying depressed individuals in online forums. In their work, the models were built to consider two levels of operation. One aspect of the model uses post-level operation, which was used to learn the representation of the post. Another operation is at the user level, where the user's emotional state has been obtained using the overall posts. Rao, Guozheng, et al. [14] proposed a method for replacing the recurrent connections typically used in recurrent neural networks with gated temporal convolutions. The gated temporal convolutions can be used in parallelizing over individual words of every user's post. This method was used on the large-scale novel Reddit Self-reported Depression Diagnosis (RSDD) dataset [15], containing over 9,000 diagnosed users with depression.

## 3. DATASET

Data is the primary source for any machine learning or deep learning algorithms to work. Dataset for our experiment was taken from CLEF eRisk: Early risk prediction on the Internet | CLEF 2022 workshop (irlab.org). The coordinators worked extensively on deidentifying the posts and labeling the users as depressed or not. The data released for this shared task contains the social media posts of 2348 users. These 2348 users wrote about 565,000 individual posts. Each user's post-collection has either been labeled as "Not depressed (0)" or "Depressed (1)". All the user posts have the following fields

1. Title
2. Date of posting
3. Text message
4. Meta Info

The dataset has been split into 3 categories (Train, Test and Dev) with ratio of 8:1:1. Table 1 describes the number of documents in each category.

Table 1: Number of documents per model building category

| **Train** | **Dev** | **Test** |
|---|---|---|
| **1878** | 235 | 235 |

Following table describes the distribution of the samples across the two different classes.

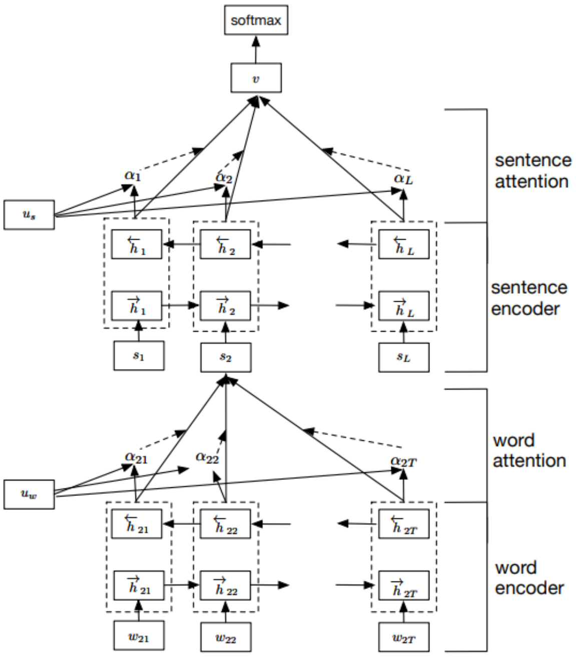
Table 2: Number of documents per category and per class

| | ***Train*** | ***Dev*** | ***Test*** |
|---|---|---|---|
| ***Not Depressed (0)*** | 1747 | 219 | 218 |
| ***Depressed (1)*** | 131 | 16 | 17 |

## 5. HIERARCHICAL ATTENTION NETWORKS FOR EARLY DETECTION OF DEPRESSION

Neural Networks are effective on text classification tasks. Hierarchical Attention Networks improve the effectiveness by considering document structure. HAN models use stacked recurrent neural networks on the word level and then an attention model to identify important words to the sentence's meaning. The representation of those informative words forms a sentence vector. A similar attention model identifies meaningful sentences for the whole document in the next layer. A vector representation of those meaningful sentences is generated. This vector representation is used for final text classification.

*HAN Architecture*



### 5.1. Experimental setup

We performed an 8:1:1 stratified split of the available dataset into train, dev, and test. Every input document contains all historical posts by the user. It is important to consider all the posts from the user for this experiment. In the pre-processing steps, we considered each message from a subject as a sentence/paragraph in the document. The sequence of messages from the same subject forms the document for the subject. For example, the first post from the user will be at the top of the document and the most recent message will be at the bottom of the document.

In the next step, we used pre-trained Glove embeddings to initialize the embeddings for the words in the corpus. The Glove embeddings were used for both Sentence level embeddings and word level embedding. Based on the architecture of the HAN models, the first attention network focuses on the words or tokens using word encoders. At the next attention network focuses on the Sentences using sentence encoders. Final output from this network produces the classification output.

We build SVM, Logistic Regression, LSTM and CNN models using the same dataset and compared the performance. Summary of the LSTM and CNN models used for comparison is given in Table 3 and Table 4.

Table 3: LSTM Model Parameters

```
__________________________________________________
__________
Layer (type)              Output Shape           Param #
========================================
===============
inputs (Input Layer)         [(None, 150)]            0
__________________________________________________
__________
embedding (Embedding)        (None, 150, 50)
500000
__________________________________________________
__________
Lstm (LSTM)                  (None, 64)             29440
__________________________________________________
__________
FC1 (Dense)                  (None, 256)            16640
__________________________________________________
__________
activation (Activation)      (None, 256)              0
__________________________________________________
__________
dropout (Dropout)            (None, 256)              0
__________________________________________________
__________
Out Layer (Dense)            (None, 1)               257
__________________________________________________
__________
activation_1 (Activation)    (None, 1)                0
========================================
===============
Total params: 546,337
Trainable params: 546,337
Non-trainable params: 0
```

Table 4: CNN Model Parameters

```
__________________________________________________
__________
Layer (type)              Output Shape           Param #
========================================
===============
input_1 (Input Layer)        [(None, None)]           0
__________________________________________________
__________
embedding (Embedding)        (None, None, 128)
2560000
__________________________________________________
__________
dropout (Dropout)            (None, None, 128)        0
__________________________________________________
__________
conv1d (Conv1D)              (None, None, 128)
114816
__________________________________________________
__________
conv1d_1 (Conv1D)            (None, None, 128)
114816
__________________________________________________
__________
global_max_pooling1d (Global (None, 128)              0
__________________________________________________
__________
dense (Dense)                (None, 128)            16512
__________________________________________________
__________
dropout_1 (Dropout)          (None, 128)              0
__________________________________________________
__________
predictions (Dense)          (None, 1)               129
========================================
===============
Total params: 2,806,273
Trainable params: 2,806,273
Non-trainable params: 0
```

### 5.2. Experiment Results

Table 5, lists the different accuracy metrics of the HAN models based on the “dev” set. The accuracy of the model is around ~97%. We will ignore this metric because the class imbalance problem. From the data set distribution, we notice that a greater number of documents belong to “non depressed” category. So, it is important to consider other metrics such as Precision, Recall and F1 scores. F1 score of the model is 82%. While the recall of the model stands at 94%, the precision is at 72%.

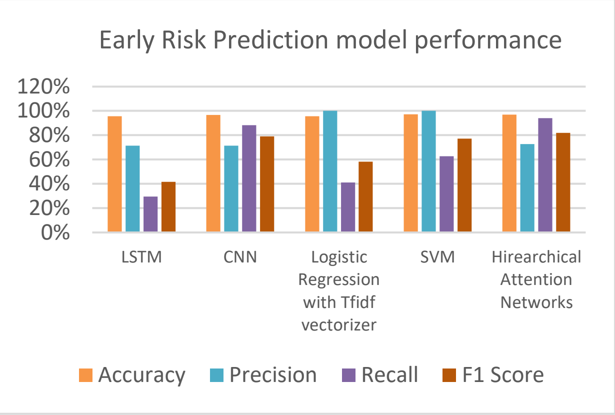
Table 5: Evaluation metrics of best model found using “dev” set

| *Metrics* | *Result* |
|---|---|
| *Accuracy* | 97% |
| *Precision* | 73% |
| *Recall* | 94% |
| *F1-score* | 82% |

Table 6 provides the performance comparison of other models with HAN models.

| *Algorithm* | *Accuracy* | *Precision* | *Recall* | *F1-Score* |
|---|---|---|---|---|
| *LSTM* | 96% | 71% | 29% | 42% |
| *CNN* | 97% | 71% | 88% | 79% |
| *Logistic Regression with TfIdf vectorizer* | 96% | 100% | 41% | 58% |
| *SVM* | 97% | 100% | 63% | 77% |
| *Hierarchical Attention Networks* | 97% | 73% | 94% | 82% |

Figure 1: A comparison of performance metrics across different model types



## 6. CONCLUSION AND FUTURE WORK

There have been several works done on identification depression using Machine Learning and Deep Learning algorithms using lexical or verbal markers. Most of the work done on this area considers linguistics aspects of the message/post by the user. The Stacked BigGRU models [14] achieved 81% F1 score. It should be noted that the Stacked BigGRU model-based approach uses the historical messages by the user along with the current posts. Hierarchical Attention network models perform. Experiment results show that for HAN the recall of the prediction is at 94%. However, the precision is at 72%. We may not consider the accuracy 97% due to the biased nature of the data set. There are more "non depressed" classes than the "depressed" classes. The F1 score is in line with some earlier studies that used Multi-Gated LeakyReLU CNN (MGL-CNN) algorithms. In comparison with other deep learning algorithms, the precision is higher for the HAN models. SVM and Logistic regression algorithms yields very high precision at the cost of recall. Recent studies shows that introduction of the Graph Neural Networks improves the classification performance. So, in order to improve the precision, GCN’s can be considered as a next step.

## REFERENCES

1. Thieme, Anja, Danielle Belgrave, and Gavin Doherty. "Machine learning in mental health: A systematic review of the HCI literature to support the development of effective and implementable ML systems." ACM Transactions on Computer-Human Interaction (TOCHI) 27.5 (2020): 1-53.
2. Lee, Yena, et al. "Applications of machine learning algorithms to predict therapeutic outcomes in depression: a meta-analysis and systematic

review." Journal of affective disorders 241 (2018): 519-532.

3. Liu, Zengjian, et al. "De-identification of clinical notes via recurrent neural network and conditional random field." Journal of biomedical informatics 75 (2017): S34-S42.
4. Durstewitz, Daniel, Georgia Koppe, and Andreas Meyer-Lindenberg. "Deep neural networks in psychiatry." Molecular psychiatry 24.11 (2019): 1583-1598.
5. Zogan, Hamad, et al. "Depressionnet: learning multi-modalities with user post summarization for depression detection on social media." Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021.
6. Zhang, Yipeng, et al. "Monitoring depression trends on Twitter during the COVID-19 pandemic: Observational study." JMIR infodemiology 1.1 (2021): e26769.
7. Orabi, Ahmed Husseini, et al. "Deep learning for depression detection of twitter users." Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic. 2018.
8. Jeffrey Pennington, Richard Socher, and Christopher D.Manning. 2014. Glove: Global vectors for word representation. In Empirical Methods in Natural Language Processing (EMNLP), pages 1532–1543.
9. Guntuku, Sharath Chandra, et al. "Understanding and measuring psychological stress using social media." Proceedings of the International AAAI Conference on Web and Social Media. Vol. 13. 2019.
10. Tuke, Curtis Murray1 Lewis Mitchell1 Simon, and Mark Mackay. "Symptom Extraction from the Narratives of Personal Experiences with COVID-19 on Reddit." (2021).
11. Trotzek, Marcel, Sven Koitka, and Christoph M. Friedrich. "Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences." IEEE Transactions on Knowledge and Data Engineering 32.3 (2018): 588-601.
12. K. Gowen, M. Deschaine, D. Gruttadara, and D. Markey, “Young adults with mental health conditions and social networking websites: Seeking tools to build community,” Psychiatric Rehabil. J., vol. 35, no. 3, pp. 245–250, 2012.
13. J. A. Naslund, S. W. Grande, K. A. Aschbrenner, and G. Elwyn, “Naturally occurring peer support through social media: the experiences of individuals with severe mental illness using youtube,” PLOS One, vol. 9, no. 10, 2014, Art. no. e110171.
14. Rao, Guozheng, et al. "MGL-CNN: A hierarchical posts representations model for identifying depressed individuals in online forums." IEEE Access 8 (2020): 32395-32403.
15. Yates, Andrew, Arman Cohan, and Nazli Goharian. "Depression and self-harm risk assessment in online forums." arXiv preprint arXiv:1709.01848 (2017).
16. Schwartz, H. Andrew, et al. "Towards assessing changes in degree of depression through facebook." Proceedings of the workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality. 2014.
17. Thompson, Paul, Craig Bryan, and Chris Poulin. "Predicting military and veteran suicide risk: Cultural aspects." Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2014.
18. Naslund, J., Aschbrenner, K., Marsch, L., & Bartels, S. (2016). The future of mental health care: Peer-to-peer support and social media. Epidemiology and Psychiatric Sciences, 25(2), 113-122. doi:10.1017/S2045796015001067
19. Harrigian, Keith, Carlos Aguirre, and Mark Dredze. "On the state of social media data for mental health research." arXiv preprint arXiv:2011.05233 (2020).
20. Zirikly, Ayah, et al. "CLPsych 2019 shared task: Predicting the degree of suicide risk in Reddit posts." Proceedings of the sixth workshop on computational linguistics and clinical psychology. 2019.
21. Amir, Silvio, Mark Dredze, and John W. Ayers. "Mental health surveillance over social media with digital cohorts." Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology. 2019.
22. Sekulić, Ivan, and Michael Strube. "Adapting deep learning methods for mental health prediction on social media." arXiv preprint arXiv:2003.07634 (2020).
23. Gamaarachchige, Prasadith Kirinde, and Diana Inkpen. "Multi-task, multi-channel, multi-input

learning for mental illness detection using social media text." Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019). 2019.

24. Turcan, Elsbeth, and Kathleen McKeown. "Dreaddit: A Reddit dataset for stress analysis in social media." arXiv preprint arXiv:1911.00133 (2019).
25. Schoene, Annika M., et al. "Dilated lstm with attention for classification of suicide notes." Proceedings of the tenth international workshop on health text mining and information analysis (LOUHI 2019). 2019.
26. Cao, Lei, et al. "Latent suicide risk detection on microblog via suicide-oriented word embeddings and layered attention." arXiv preprint arXiv:1910.12038 (2019).
27. Matero, Matthew, et al. "Suicide risk assessment with multi-level dual-context language and BERT." Proceedings of the sixth workshop on computational linguistics and clinical psychology. 2019.
28. Tadesse, Michael M., et al. "Detection of depression-related posts in reddit social media forum." IEEE Access 7 (2019): 44883-44893.
29. Franz, Peter J., et al. "Using topic modeling to detect and describe self-injurious and related content on a large-scale digital platform." Suicide and Life-Threatening Behavior 50.1 (2020): 5-18.
30. Malmasi, Shervin, Marcos Zampieri, and Mark Dras. "Predicting post severity in mental health forums." Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology. 2016.
31. Brew, Chris. "Classifying ReachOut posts with a radial basis function SVM." Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology. 2016.
32. Yang, Zichao, et al. "Hierarchical attention networks for document classification." Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies. 2016.
33. Losada, David E., Fabio Crestani, and Javier Parapar. "eRISK 2017: CLEF lab on early risk prediction on the internet: experimental foundations." International Conference of the Cross-Language Evaluation Forum for European Languages. Springer, Cham, 2017.